Beat Tracking

(or, can the computer clap along with the music?)

Robin D. Morris, RIACS, NASA Ames Research Center
William A. Sethares, Dept of E&CE, University of Wisconsin-Madison

We present a method of tracking the beat in musical performances. The raw audio is first processed into a collection of ``rhythm tracks'', which attempt to make more explicit the rhythmic structure of the audio. Each of these rhythm tracks is considered as an (approximately) independent observation of the underlying rhythmic structure of the audio. A simple generative model for the structure of the rhythm tracks is hypothesised, and a sequential Monte Carlo algorithm is developed for tracking the parameters of the model through time. The output of the algorithm are parameters that represent when the first beat occurs, the interval between beats, and the rate of change of the beat period. The behavior of the algorithm is investigated on a variety of musical sources. The technique can be applied directly to a digitized performance (i.e., a soundfile) and does not require a musical score or a MIDI transcription, though an analogous method can also be applied when such extra information is available. Several examples are presented that highlight both the strengths and weaknesses of the approach.

Introduction

- A large proportion of the most “significant audio events” occur on or near the beat.
- Events which are less significant rhythmically tend to occupy the spaces between the beats

Hypothesis

- Times near the beat will have significant energy, while times between the beat will be unenergetic.

Rhythm Tracks

- We propose four methods for extracting rhythmically important structure from the audio track
- Because they look at different characteristics of the audio signal, each off which is related to the rhythmic structure, we consider them as (quasi-)independent observations.
- The rhythm tracks have a sampling rate one hundred times less than the original audio track, as the beat period tends to be on the order of one second.

Rhythm Track 1: Time domain energy method

- Appropriate when the envelope of the sound waveform displays the beat
- Group the sampled data into m overlapping sets, each containing N samples.
- Let $x_n[k]$ represent the k^{th} element in the n^{th} set.
- Energy in the n^{th} set is $e[n]=\sum_{k=1}^{N} x_n^2[k]$
- Rhythm track $r_E[n]=\frac{\partial}{\partial t}e[n]$

Rhythm Track 2: Spectral center

- Locates the mean (center of mass) of the spectrum;

the frequency where $$\int_{f=0}^{fc} X_n^2(f)df = \int_{fc}^{\infty} X_n^2(f)df$$

(where $\mathrm{x_n}$ is the nth set of N (overlapping samples), and $\mathrm{X_n}$ is the FFT of $\mathrm{x_n}$)

- Rhythm track $\mathrm{r_{cm}[n]}$ is the rate of change of $\mathrm{f_c}$.
- Sensitive to pitch changes and changes in energy distribution.
- Insensitive to amplitude.

Rhythm Track 3: Spectral dispersion

- Measure the spread of the spectrum about its mean

$$sd[n] = \int_{f=0}^{\infty} X_n^2(f)\left|f - f_c\right| df$$

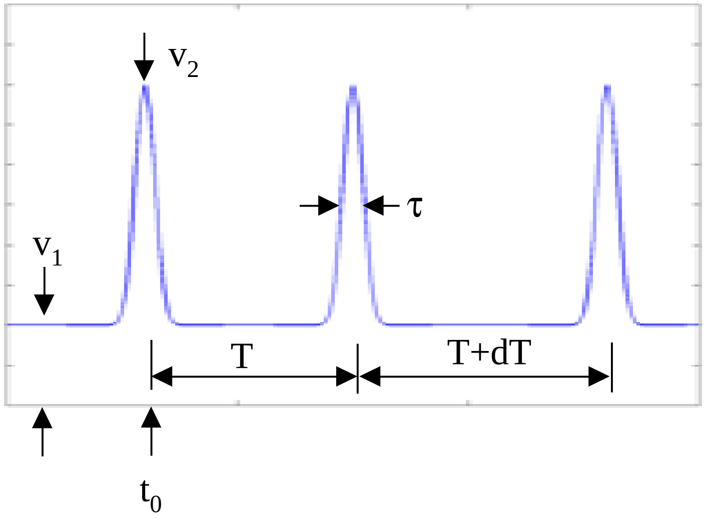
- Crude measure of the distribution of energy in the spectrum
  - small values: energy concentrated near centre
  - large values: energy widely dispersed
- Rhythm track r_{sd} is the rate of change of $\mathrm{sd[n]}$.

Rhythm Track 4: Group delay

- The unwrapped phase of X_n lies close to a straight line for many musical tracks.
- The slope, t_g, of this line defines the group delay method, $r_{tg}[n]$.
- t_g can be shown to give an estimate of where the energy is concentrated.
- It is insensitive to amplitude.

A Generative Model for the Rhythm Tracks

- We propose the simplest (and probably over-simplified) generative model for the rhythm tracks.



The model has two sets of parameters, those describing the structure, and those describing the timing.

The structural parameters are:

- “off the beat” variance, v_1
- “on the beat” variance, v_2
- “beat width”, τ

These parameters have been found to remain fairly constant across different musical pieces

Generative Model (cont.)

The timing parameters are:

- t_0 - time of the first beat
- T - beat period
- dT - rate of change of beat period

and it is these parameters that we wish to track through the audio signal

- The signal is assumed to be uncorrelated Gaussian noise, with variance defined by the amplitude of the pulse train.
  - This is over-simplified, and ignores a great deal of the structure in the rhythm tracks.
  - However, it was found to be surprisingly effective for a wide range of audio signals.

Tracking Using Particle Filters

- p(t_0,T,dT | current block of rhythm tracks) $\propto$ p(current block of rhythm tracks | t_0,T,dT) p(t_0,T,dT | previous blocks of rhythm tracks)
- Because the tracks are considered to be independent, the posterior is

  p(rhythm track 1 | t_0,T,dT) x

  p(rhythm track 2 | t_0,T,dT) x

  :

  p(t_0,T,dT | previous blocks of rhythm tracks)
- and p(rhythm track | t_0,T,dT) is modeled as a product of Gaussians with the structured pattern of variances described above.

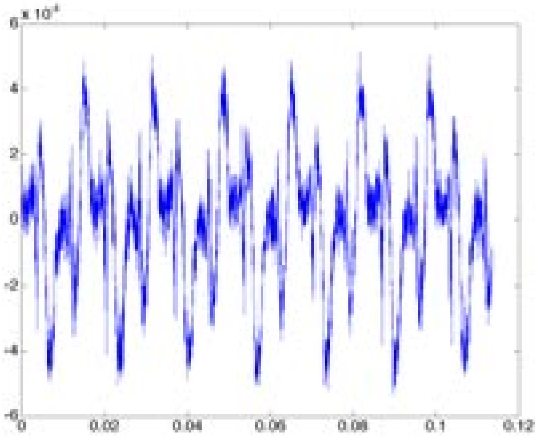
Tracking Using Particle Filters (cont.)

- Prediction - a simple Gaussian diffusion is added to the samples
- Update - the likelihood discussed above is used to re-weight the samples prior to re-sampling.

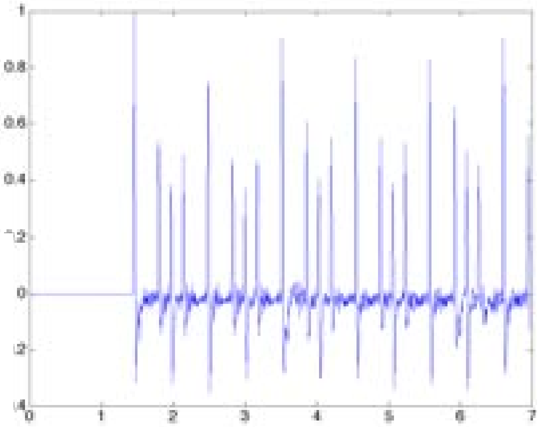
Results

- We present results for four audio tracks:
  - “threetwo” - a track generated by a drum machine with a 3/2 beat
  - “jit” -
  - “kronos” - an extract of a track by the Kronos string quartet
  - “ska” - a ska track with more complex rhythmic structure
- For all track we used the same values for the structural parameters, and the same initial prior distribution over the timing parameters.
- Rhythm track 2 was not used.
- The original audio tracks and the tracks with the extracted beat superimposed are available for your listening pleasure.

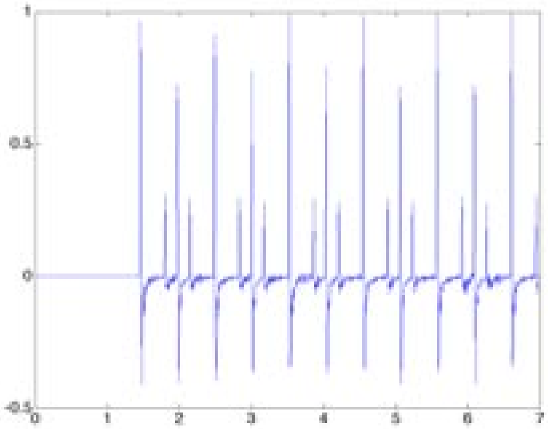
Results - “threetwo”



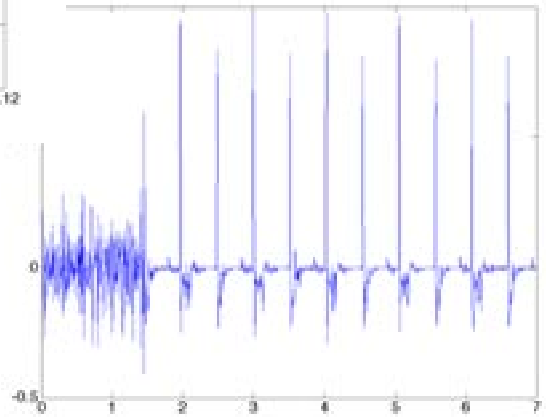
signal



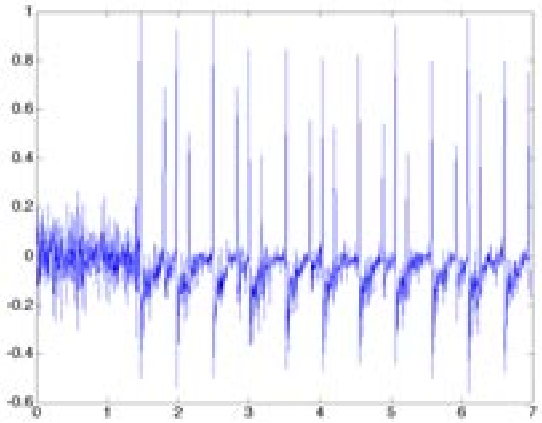
rhythm track 1



rhythm track 3

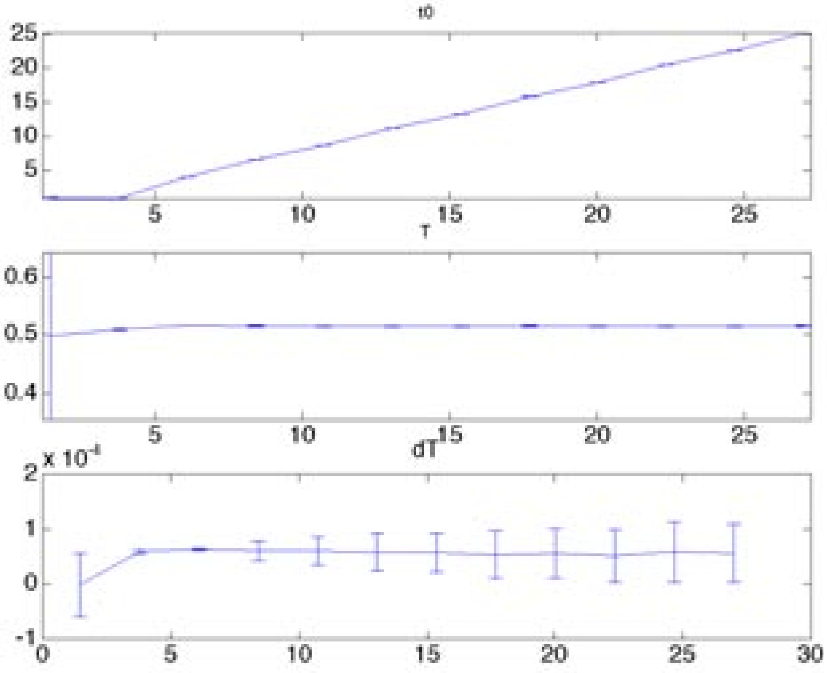


rhythm track 2



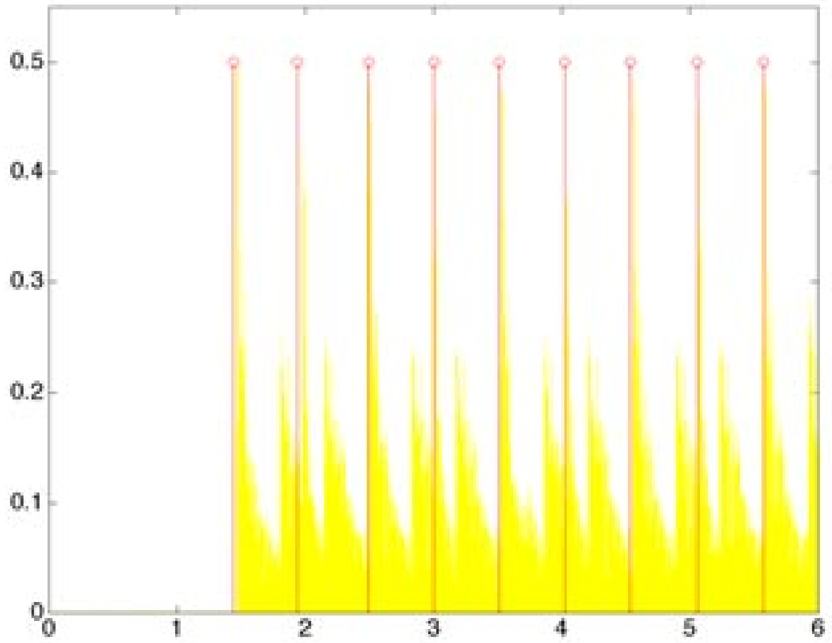
rhythm track 4

“threetwo” (cont.)

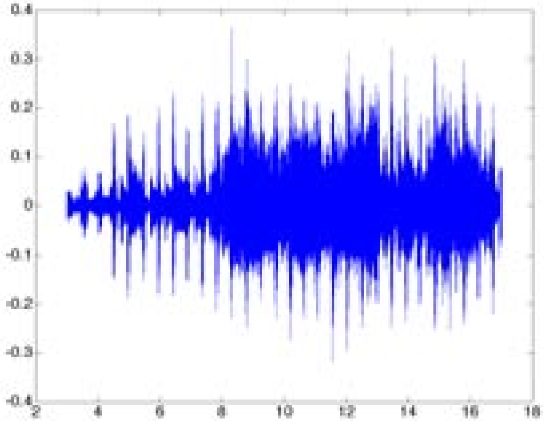


tracking results

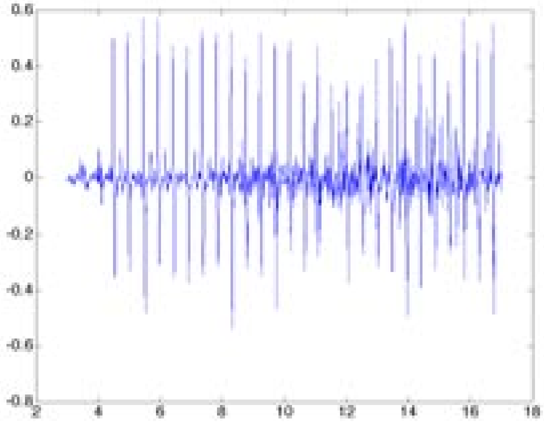
superimposed beats



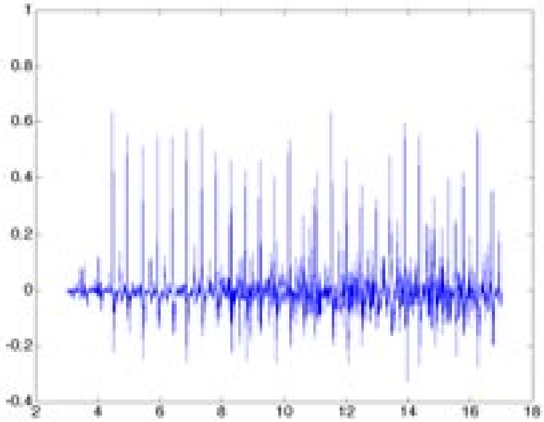
Results - "jit"



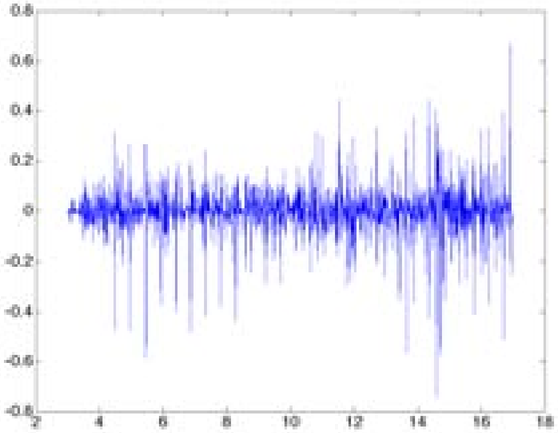
signal



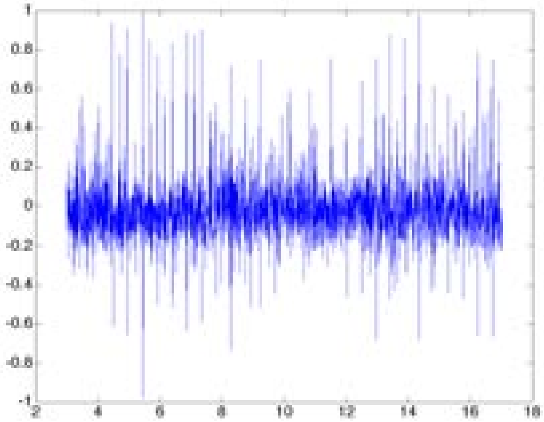
rhythm track 1



rhythm track 3

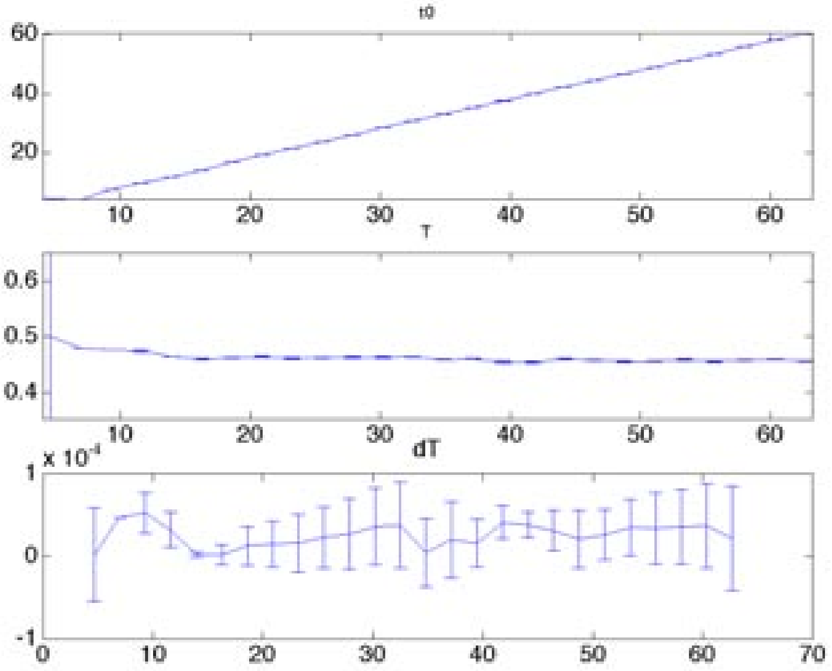


rhythm track 2



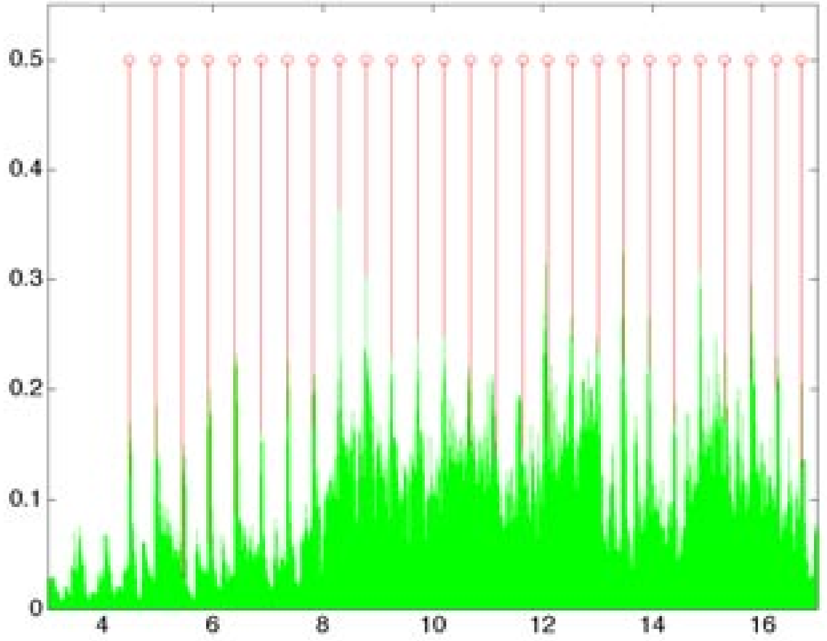
rhythm track 4

“jit” (cont.)

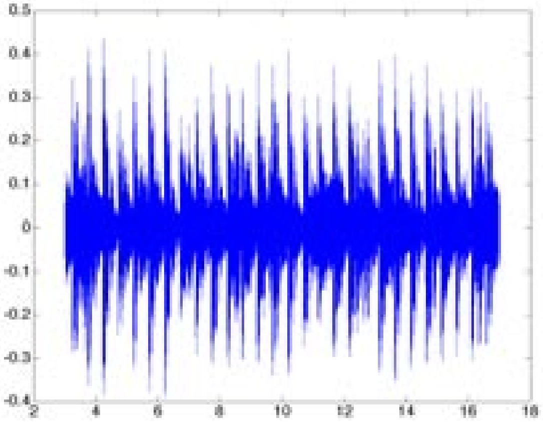


tracking results

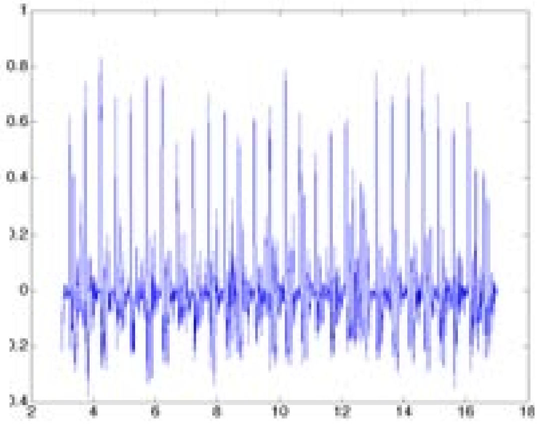
superimposed beats



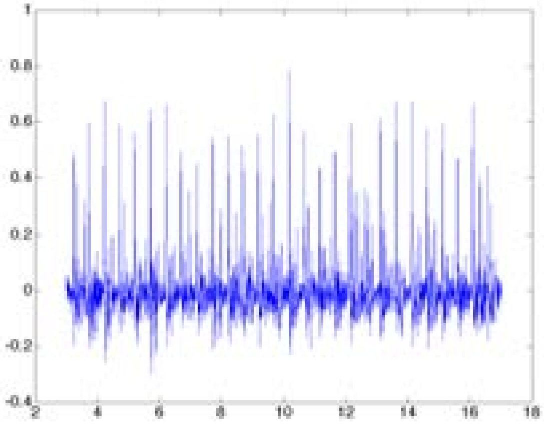
Results - “kronos”



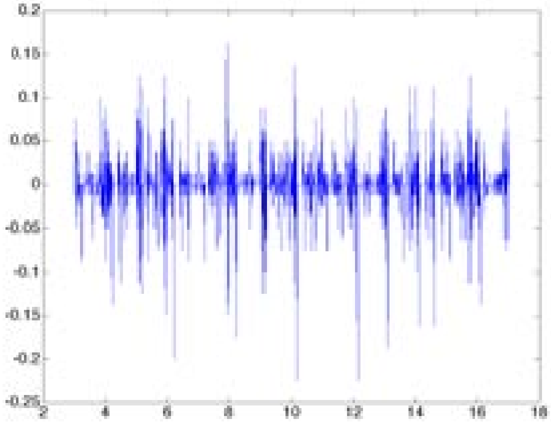
signal



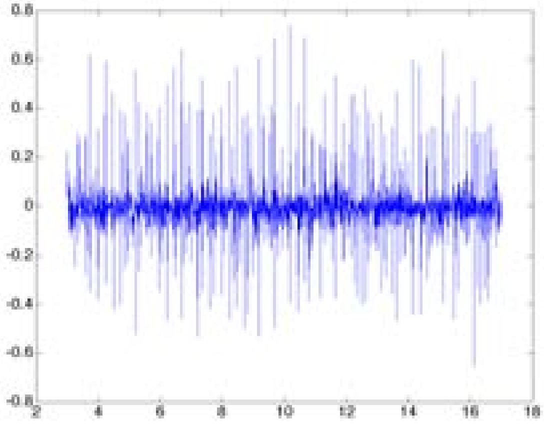
rhythm track 1



rhythm track 3

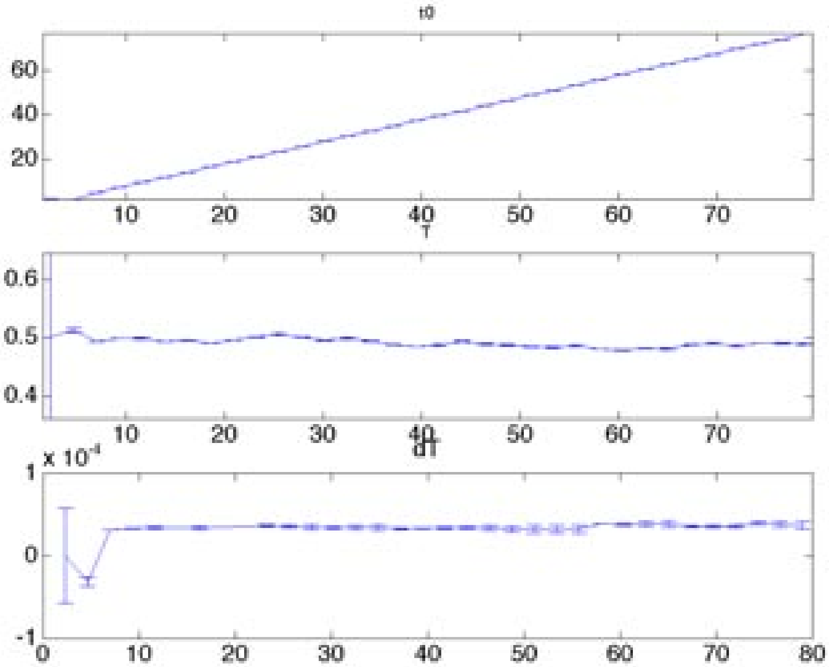


rhythm track 2



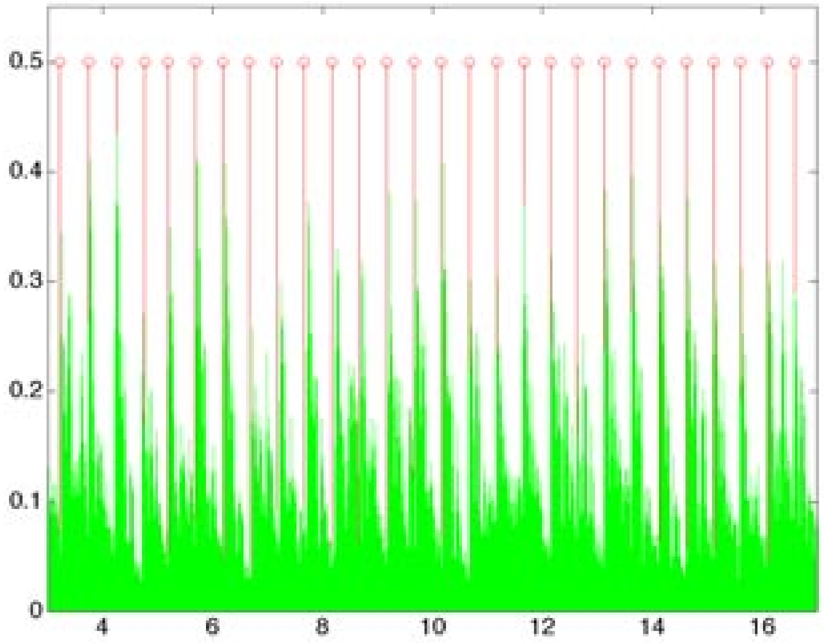
rhythm track 4

“kronos” (cont.)

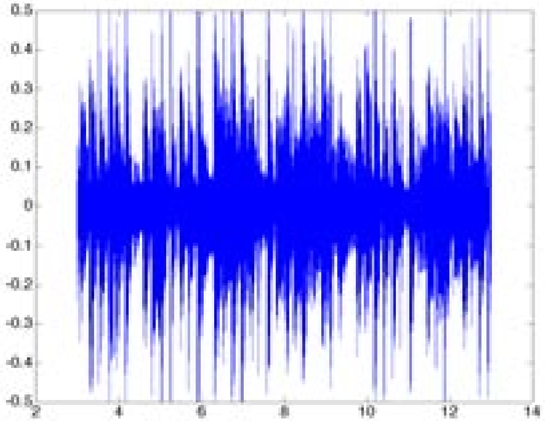


tracking results

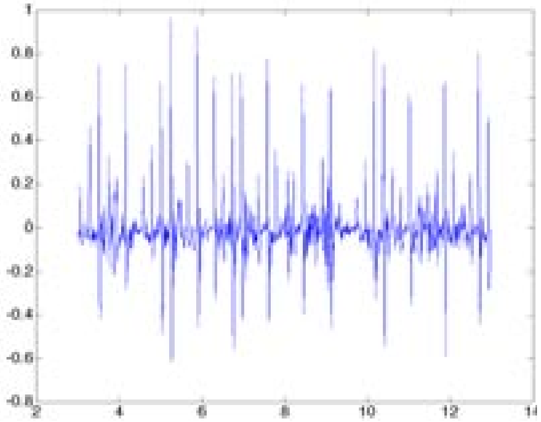
superimposed beats



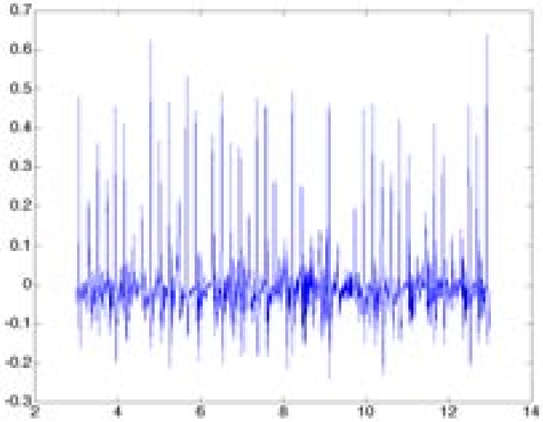
Results - “ska”



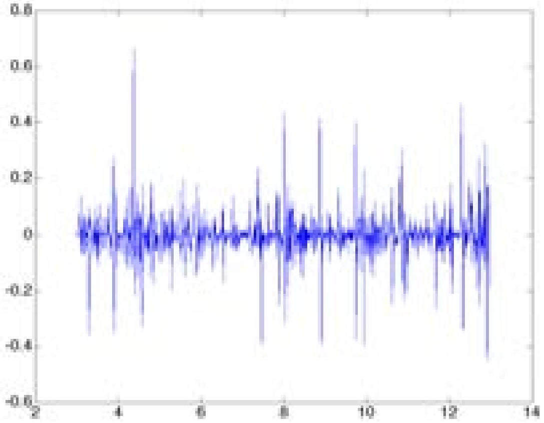
signal



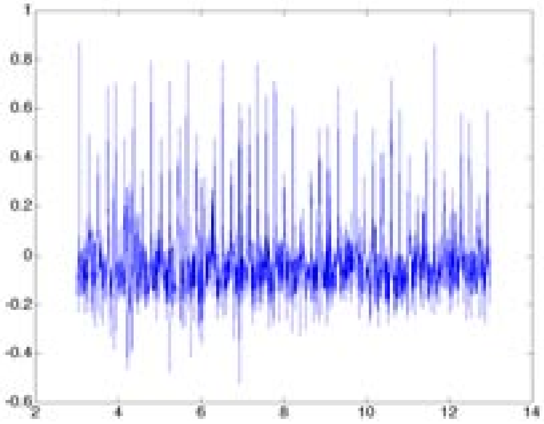
rhythm track 1



rhythm track 3

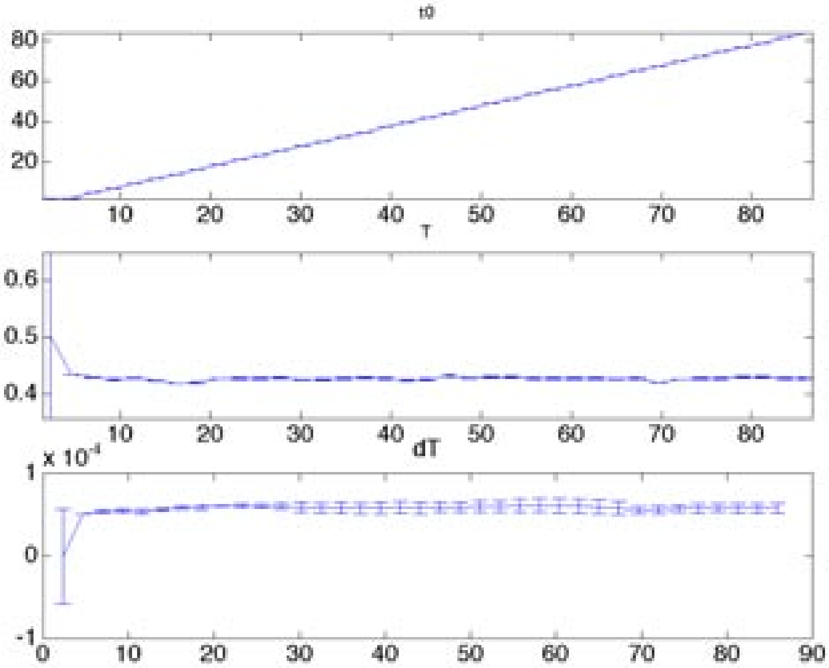


rhythm track 2



rhythm track 4

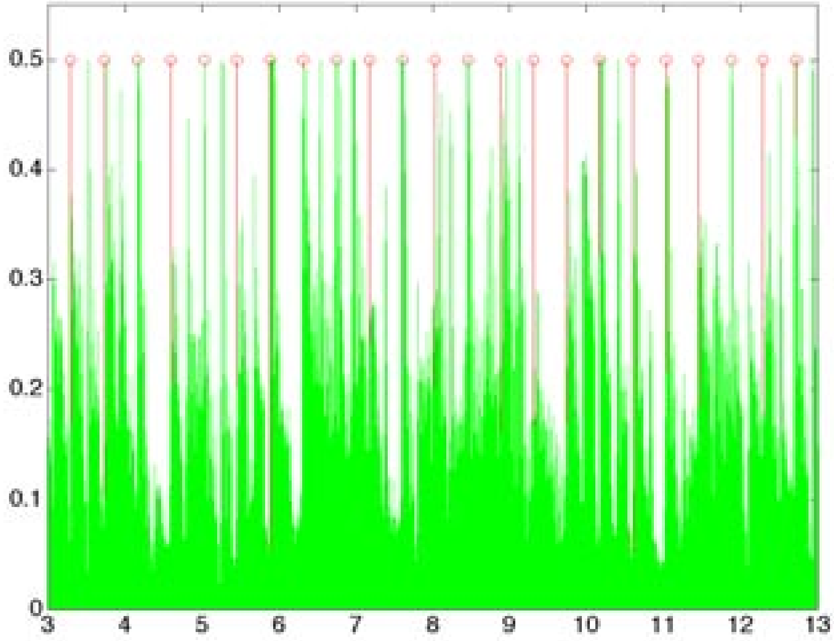
“ska” (cont.)



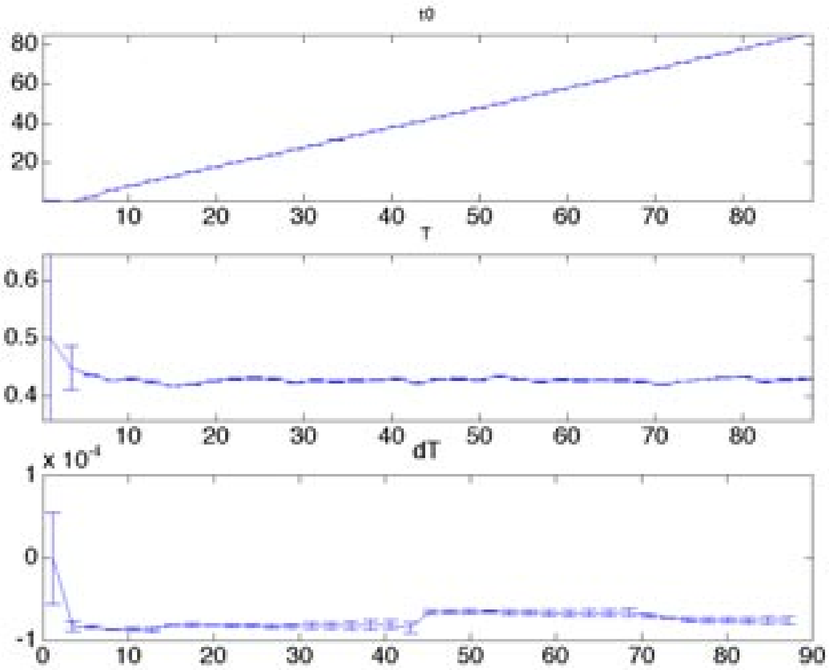
tracking results

tracking the “odd” beats

superimposed beats



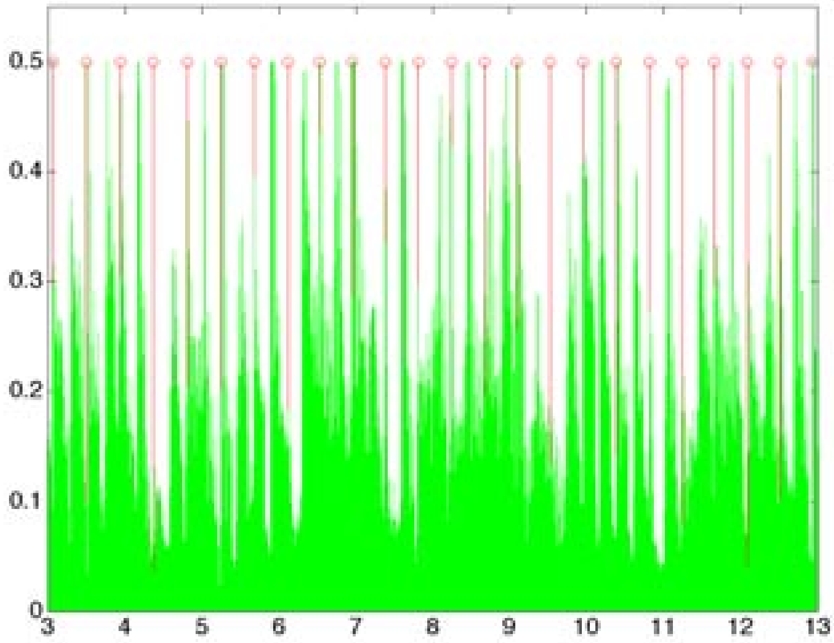
“ska” - cont.



tracking results

tracking the “even” beats

superimposed beats



Discussion

- We have presented an algorithm that successfully extracts the rhythmic structure from audio tracks.
- It has been shown to be relatively insensitive to the setting of the structural parameters.
- It has been shown to exhibit similar ambiguities to human listeners in the case where there is a definite on/off beat.
- Further testing is needed to explore the limits of the algorithm, and to compare it in a principled way with beat extraction from the same tracks by human listeners.